P 114 ✓

TABLES AND AN ALGORITHM FOR CALCULATING FUNCTIONAL GROUPS OF ORGANIC MOLECULES IN HIGH RESOLUTION MASS SPECTROMETRY

Joshua Lederberg
Professor of Genetics
School of Medicine
Stanford University

STAR (Scientific and Technical Aerospace Reports)

No. N64-21426

Submitted April 1, 1964 to NASA as an interim report under Grant No. NsG 81-60, "Cytochemical studies of planetary microorganisms".

TABLES AND AN ALGORITHM FOR CALCULATING FUNCTIONAL GROUPS OF ORGANIC MOLECULES IN HIGH RESOLUTION MASS SPECTROMETRY

Joshua Lederberg
Stanford University

Introduction

The pioneering work of Beynon[1] has demonstrated the salient power of high resolution mass spectrometry for the structural analysis of organic molecules. Instruments with resolution capabilities of 1 part in 10,000 to 100,000 are becoming commercially available. There remains the problem of computing the molecular composition from the experimentally found mass. Tables for this purpose have been published[2] but are necessarily limited in scope by the sheer number of possible combinations.

On the atomic mass scale $^{12}C = 12.00000$, the fractional mass is attributable exclusively to the non-carbon part of a molecule, so these parts can be tabulated separately and much more compactly than complete tables of composition. Such tables have been computed over the range of H from 0 to 124, N from 0 to 6 and O from 0 to 11, and published elsewhere[3] to make them generally accessible.

An alternative approach was also generated, which gives even more compact tables and an algorithm more suitable for computer-oriented analysis, and is presented here.

The CH_2=14 Algorithm

Where molecules in homologous series are in question, or the investigator has a definite combination of functional groups in mind, the CH_2=14 algorithm will more than justify some additional arithmetic on the part of the specialized user. Furthermore, its tables are even more compact: roughly speaking, where an independent variable of the C=12 system is the number of H atoms, plausibly from 0 to 124, in the CH_2=14 system we use degrees of unsaturation from 0 to 30. This permits us to encompass most molecules and radicals, N from 0 to 8 and O from 0 to 13 in Table 2. In addition, the entries appear in an order related to the complexity of the molecule. The topmost, simpler areas of the tables may become quite familiar with extensive use, since e.g. alkanes, monocarboxylic acids, dicarboxylic acids, monoketones, etc. each have a singular location in the tables.

The logic of extracting the formula by division and examining the remainder is similar to that of the C=12 system. However, as Kendrick[4] has pointed out, using CH_2 as a base relates the formula to the structural concept of a fundamental hydrocarbon, with functional substituents. Instead of taking C, H, O, N as the variables, we take a formula as Terminal H +

CH_2, C:, O, NH

parts, i.e. saturated carbons, double bonds, oxygen, and amino-functions. For computational purposes, -N= is regarded as -NH- +C:, $-CH_2-CH=N-$ is treated as $\overset{H}{-C}=\overset{H}{C}-NH-$. This leads to a mathematical fiction in treating molecules which have more double bonds than C atoms. "Double bonds" includes rings (equivalent to one double bond each) and -C=C- functions (equivalent to two C:), i.e. C: can be read as "degrees of unsaturation" together with an equal number of C atoms.

In principle we can implement this scheme of extracting the CH_2 part by shifting to a mass scale $^{12}CH_2$ = 14.00000, and then dividing by 14. This would be done by multiplying the found mass by 0.99888337 = (14.00000/14.01565). We would then tabulate the <u>mass defect</u>, the amount by which the decimal value falls short of an integral mass number. There will be a characteristic defect on this scale for every class of organic compound. A few classes of compounds exemplified by ethane, ethyl radical, and ethylamine have a mass excess, i.e. negative defect, -.01339, -.00669 and -.00753 respectively; i.e. their masses are 30.01339, 29.00669 and 43.00753, respectively, when expressed on the CH_2=14 scale, while higher monologues with n additional CH_2 groups will be precisely 14n larger. In these cases their functional groups fail to overbalance the mass excess of the two terminal hydrogens of the basic structure. Alkenes, C_nH_{2n}, will have a defect of zero.

Other functional groups will have equally characteristic mass defects, simply the sum of the contribution of each part (see Table 3).

In actual computation the multiplier .99888 is clumsy and inefficient. Where m = exact mass on ^{12}scale, it is better to take m (0.99888337) = m-.00111663m, especially as the factor can usually be truncated, and only the decimal corrections need be kept in hand. This will be readily seen in working through the examples. Table 1 calculates the correction from the integer part of the mass (with reasonable precision for most purposes) so the only arithmetic actually needed is to subtract the entry of Table 1 from the observed mass, and use the difference in scanning the main table (Table 2).

Terminal H numbers 2 for complete molecules, 1 for radicals, the latter being signified by * in the table.

<u>Example.</u> An alkaloid analysed to contain about N_4 and O_6 to O_8 returned a mass reading of 718.37430 ± .00600.

	Quotient		Residue		Decimal
1. From Table 1, 718.37430 =	51 (x 14)	+	4	+	.37430

2. Calculate m (1-0.00111663). From Table 1:

718	⟶	80174*		
.37430	⟶	37430		
Defect		42744	±	600

(*The table gives the value for the integer 718. A more precise measure of the defect by interpolation or direct multiplication would be 80216 - 37430 = 42786; with practice, the interpolation can be done by eye if necessary.)

3. From Table 2, integer residue class 4, the following entries are within the range 42144 - 43344 and are intact molecules (not*):

Defect	C:	NH	O	$\equiv(CH_2)_x$
42192	17	8	7	31
42325	12	4	11	27
42378	30	6	0	32
42511	25	2	4	28
42727	8	6	13	28
42779	26	8	2	33
42912	21	4	6	29
43045	16	0	10	25
43314	17	6	8	30

Of these, only 42912 satisfies the compositional requirements with respect to N and O. It may be interpreted as follows:

$$2H + 21C: + 4NH + 6O + (51-29) = 22CH_2$$

or $C_{43}H_{50}N_4O_6$.

Extensions

The proper extent of the tables is a compromise between the bulk and cost of a larger table as against the effort of additional arithmetic. The most frequent extension may be in the range of NH, here limited to 8. If so indicated, 9NH can be extracted from the experimental mass by subtracting 135.09809; the entries in the table will then correspond to NH from 9 to 17 in place of 0 to 8.

Further extended versions of the tables can be computed if their utility warrants. Alternatively, other special cases can be handled with the present tables with some additional arithmetic. If additional atoms are suspected, their weights should simply be subtracted from the mass numbers found. E.g., if monochloro compounds are in question, the corresponding dichloro radicals are formed by subtracting ^{35}Cl or ^{37}Cl.

The tables are presented to 5 significant figures, which is an optimistic projection of instrumental capacity. The 5th digit is subject to rounding error of computation. The values refer to the mass of the neutral molecule rather than the positive ion, as the mass of neutral reference molecules is usually set down in calibrating the mass spectrometer.

Constants

The constants used in the computation are (on the ^{12}C scale): H = 1.00782522, N = 14.003074, O = 15.994915. They are taken from the IUPAC report[5].

Acknowledgments

These tables represent an exercise in the application of computers to biochemical problems. Research connected with this program has been supported by grants from the National Aeronautics and Space Administration (NsG 81-60), National Science Foundation (NSF G-6411), and National Institutes of Health (NB-04270-01 and 02, and AI-5160-06). I am grateful to Professor Carl Djerassi for having challenged the computer to generate the extended tables, to which the present algorithms are a rebuttal.

The programs were run under the Subalgol monitor on the IBM 7090 at Stanford University Computation Center, whose assistance to academic research is supported by an NSF grant (NSF-GP948). I am indebted to the staff of the Computation Center for their unstinting cooperation and to Mrs. Margaret Wightman for skilled and loyal assistance.

The actual operation of these programs required about one second of main frame computer time per page of tabular output.

References

1. J. H. Beynon, Mass Spectrometry and Its Applications to Organic Chemistry, Elsevier, Amsterdam, 1960.

2. J. H. Beynon and A. E. Williams, Mass and Abundance Tables for Use in Mass Spectrometry, Elsevier, Amsterdam, 1963.

3. J. Lederberg, The Computation of Molecular Formulas for Mass Spectrometry, Holden-Day, San Francisco, 1964.

4. Prior to the completion of this report a similar proposal was enunciated by E. Kendrick, A mass scale based on CH_2 = 14.0000 for high resolution mass spectrometry of organic compounds, Analytical Chemistry 35:2146, 1963. His implementation, oriented towards petroleum constituents, is substantially different, but the underlying concept closely anticipates the present scheme.

5. A. E. Cameron and E. Wichers, Report of the International Commission on Atomic Weights (1961), Journal of the American Chemical Society, 84: 4175, 1962.

BALGOL PROGRAM

For Computation of Tables of Mass Defects by Composition

```
2   MIN3000   LEDERBERG  MASSCORRECTIONS..14 ALGORITHM
               STANFORD UNIVERSITY COMPILER -- VERSION OF  1/27/64

144...     COMMENT*** OUTPUT ON TAPE       $

      *-SYMBOL

      *-SPACE
144...     COMMENT*** OUTPUT ON TAPE 648$
144...     INTEGER OTHERWISE$
144...     GLOBAL INTEGER COLUMNS$
144...     ARRAY A(2000,10)$
144... PROCEDURE COLOUT(LINES,COLS $ VAR,V1,V2,V3 $ OUT,FORM) $
156...     BEGIN INTEGER OTHERWISE $ SL = V2.LINES $
161...     M = (COLS-1)SL + V1 $ COLUMNS = COLS $ SM = SL + V1 - V2 $
176...     COMMENT WRITES MIN( LINES.COLS , (V3+V2-V1)/V2 ) OUT'S $
207...     FOR I = (V1,V2,SM) $ BEGIN
213...        UNTIL M LEQ V3 $ (M = M-SL $ COLUMNS = COLUMNS-1) $
224...        WRITE($$0, FORM) $ M = M + V2 END $
233...     RETURN $ OUTPUT O(FOR VAR=(I,SL,M) $ OUT())
256...     END COLOUT() $
304... PROCEDURE COLIST(COLS $ I,I1,I2,I3 $ OUT, FORM) $
316...     BEGIN INTEGER OTHERWISE $
316...     J = 60 COLS.I2 $
322...     FOR K = (I1, J, I3) $ COLOUT(60,COLS $ I,K,I2,I3 $ OUT,FORM) $
351...     RETURN END COLIST() $
372... PROCEDURE SORTS(FILL,N , VALUES() , KEY()) $
      *-PRINT
      *PRINT
664...    FOR I=(1,1,1000)$A(I,9)=I$
705...        FOR M=(0,1,13)$ BEGIN
742...     FUNCTION CR(D)=134000D + 58626NH +2294540 - 67000HR$
751...     I=0$
752...     FOR D=(0,1,30)$
763...     FOR NH =(0,1,8)$
776...     (( HR=-MOD(M+NH,2)$
1011...    ( O = MOD( M + 1398 + 2D - HR - NH,14)/2 ) ) $
1025...    FOR O=O,O+7$(
1053...    SUM = 12D + 15 NH + 16O + 2 + HR$
1055...    J=1$ I=I+1$ FOR A(I,J)=CR(D),D,NH,O,HR,SUM/14$ J=J+1))$
1162...    ILIM=I$
1164...  SORTS(0,ILIM,A(,9),A(,1))$
1177...    ILIM = MIN(ILIM,540)$
1204...    COLIST(3$ K,1,1,ILIM$ SOME,NUMBERS)$
1214...    END$
1215...      OUTPUT SOME(FOR I=A(K,9)$ FOR HR=A(I,5) $((A(I,1)-133950)/100,
1270...                 FOR J =    2,3,4,6 $  A(I,J)))$
1302...    FORMAT NUMBERS($COLUMNS$(B6,L5,B,2I3,I4,B,$-HR$('*'),$1+HR$(' '),
1341...   I3),W)$
1341...    FINISH$
```

NAME	FUNCTIONAL GROUP	CONTRIBUTION[1] TO INTEGER RESIDUE	CONTRIBUTION[1] TO MASS DEFECT	BASE COMPOUND	MASS $^{12}C=12$ SCALE	RESIDUE CLASS	MASS DEFECT OF COMPOUND
-ANE	2H-			H_2	2.01565	2	-.01339
-ANYL (RADICAL)	H-	-1	.00669	H-	1.0078252	1	.00669
-AMINE	NH	1	.00586	NH_3	17.026549	3	-.00753
-OL or ETHER	O	2	.02294	H_2O	18.010565	4	.00955
-AL or -ONE or -ENOL	C=O	0	.03634	CH_2O	30.010565	2	.02295
-ENE or RING	-C=	-2	.01339	$CH_2=CH_2$	28.03130	0	.00000
-OIC ACID	COO	2	.05928	HCOOH	46.005480	4	.04589
AMIDE	CONH	1	.04220	$HCONH_2$	45.021464	3	.02881
-THIOL or THIOETHER	S	4	.06363	H_2S	33.98772	6	.05023
AMINO ACID	NHCOO	3	.06514	NH_2COOH	61.016379	5	.05175
-DIOL	2O	4	.04588	H_2O_2	34.005480	6	.03249
-TRIOL	3O	6	.06883	$C_3H_5(OH)_3$	92.047345	8	.05544
BENZENE	4C=	-8	.05359	C_6H_6	42.046950	8	.04020
HYDROXY ACID	OCOO	4	.08223	H_2CO_3	62.000395	6	.06884
PHOSPHORIC ESTER	OPO_3	12	.14586	H_3PO_4	97.97693	0	.13247
SULFONIC ACID	SO_3	10	.13246	H_2SO_3	81.97246	12	.11907
SULFATE ESTER	OSO_3	12	.15540	H_2SO_4	97.96738	0	.14201
NITRILE	CN	11	.02595	HCN	27.01090	12	.01926
CYANATE	CNO	13	.04890	HCNO	43.00581	0	.04221
THIOCYANATE	CNS	1	.08958	HCNS	58.98297	2	.08289
CHLORIDE	35 Cl	6	.07018	HCl	35.97668	7	.06349
CHLORIDE	37 Cl	8	.07537	HCl	37.97372	9	.06868
BROMIDE	79 Br	8	.16977	HBr	79.92617	9	.16308
BROMIDE	81 Br	10	.17401	HBr	81.92416	11	.16732
FLUORINE	F	5	.02283	HF	20.0062	6	.01614

Table 3. Mass Defects (CH_2=14 scale) of Simple Functions

[1]as additive perturbation of a compound previously calculated.